The Thesis Committee for Zihan Yang
Certifies that this is the approved version of the following Thesis:

This Video is Sponsored: Text and Sentiment Analysis of YouTube Health-related Vlog Comments and Brand Endorsement Effectiveness

APPROVED BY
SUPERVISING COMMITTEE:

Gary Wilcox, Supervisor

Natalie Brown-Devlin

This Video is Sponsored: Text and Sentiment Analysis of YouTube Health-related Vlog Comments and Brand Endorsement Effectiveness

by

Zihan Yang

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

December 2020

Acknowledgements

I would like to express the deepest gratitude to my parents and friends. Their selfless love and support encourage me to finish this work during such a difficult time worldwide.

I would like to thank my supervising committee members, Dr. Gary Wilcox and Dr. Natalie Brown-Devlin, for their relentless support. It has been my fortune to learn and work with such experienced and fantastic professors.

I would like to express my appreciation to the faculty and staff of the Stan Richards School of Advertising and Public Relations, as well as the faculty of the Department of Statistic and Data Science and the College of Education. Thank them all for the advice, knowledge, care and inspiration in the past two and a half years. It was truly a memorable period to study at the University of Texas at Austin.

Abstract

This Video is Sponsored: Text and Sentiment Analysis of YouTube Health-related Vlog Comments and Brand Endorsement Effectiveness

Zihan Yang, M.A.

The University of Texas at Austin, 2020

Supervisor: Gary Wilcox

With the rapid development of YouTube and social media influencers, influencer brand endorsements have received much industry and scholarly attention. For brand endorsements made by YouTube influencers, comments under videos provide a cue for measuring endorsement effectiveness. This thesis examines the comments of both sponsored and non-sponsored YouTube health-related vlogs, as well as consumers' commenting behaviors under the vlogs and purchase intentions toward the endorsed products. Text and sentiment analysis and survey techniques are used to examine the linguistic style of comments, the commenting behaviors of consumers, and their purchase intentions. The analysis reveals that a narrative, external, and positive linguistic style is found in the comments of health-related vlogs, and consumers' positive commenting behavior leads to higher purchase intentions of healthy products.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

YouTube was founded in 2005 and has now become the world's largest online video platform, providing video content generated by users. As the third most visited website in the world, YouTube has become an important platform for advertisers to publish their ads (Zote, 2020). Because of YouTube's great online environment for individual users, there are more and more influencers emerging on the site, making the video-sharing platform one of the most suitable social platforms for influencers and influencer marketing. Given its high-quality UGC (user-generated content), companies have been considering influencers' videos as the marketing channel for their brands and products.

Among all the UGC videos on YouTube, vlog (short for video blog) is a new-generating video format with high involvement of publishers' daily lives. As the video format of blogs, vlogs provide content that mainly discusses vloggers' personal matters (Christian, 2009), which satisfies audiences' desire for imitating. The realism of vlogs makes it easier to attract audiences' attention and influences such audiences to copy vloggers' actions. Therefore, vlogs have become a preferred channel for corporate brand endorsements.

When a brand decides to cooperate with and sponsor a vlogger's videos, the vlogger is offered products, discounts, gift cards, or other services (Munnukka, 2019). The most common goal of brand endorsement by vloggers is increasing brand awareness and product sales. Therefore, after the cooperation has been built, a unique promo code is given to each vlogger. Vloggers apply the code to the "Show More" section of the product-promoting vlog. This approach is an efficient way driver of increase in sales (Loren, 2016).

The effectiveness of brand awareness and later-on purchase intentions are, however, not easy to measure. Promo codes are attractive for many audiences, but to make a decision, the late majorities and laggards in the Diffusion of Innovations are more likely to combine more information (Rogers, 2003) such as the credibility of vloggers, the comment sentiments of the video, and the review and attitude of products. Taking these considerations into account is therefore crucial.

With the growth of personal health habits pursuit, health-related vlogs including a healthy diet, fitness, and exercise have become a large category attracting a majority of people. Many vloggers and influencers in this field have millions of followers, and the top of them even sell their own products. In this context, the health-related vlog is a great representation of YouTube vlogs and brand endorsement. How do audiences' purchase intentions unfold? Do attitudes toward products, vloggers, and audience participation influence purchase intentions? The answers to these questions could offer insight into companies' influencer marketing decisions.

There are numerous studies about how audiences' attitude towards influencers or vloggers affect the effectiveness of brand endorsements, or how the interaction pattern between audience and influencers affect their attitude towards brands (Paula & Antoniya, 2017; Labrecque, 2014; Schouten, 2019; Munnukka, 2019). Further, while many studies have used text analysis to investigate the linguistic style of social media content (Younis, 2015; Asghar, 2015; Kabir, 2018), only a few of these have examined the relationship between the comments' linguistic style and the brand endorsement effectiveness.

This study, therefore, examines audiences' purchase intentions (brand endorsement effectiveness), attitudes on vloggers and brands, and the audience participation of the vlog, especially their commenting behaviors. The current study focuses on health-related vlogs and analyzes vloggers by focusing on the channels, videos, comments, and personal impressions that prompt higher measures of brand endorsement effectiveness. Two studies are contained in this paper, the first one uses the API tool Netlytic to obtain comments data of YouTube vlogs, and text analysis tool LIWC2015 is applied to analyze the linguistic style of each video's comments. The second study explores the relationship between audiences' commenting behaviors and their purchase intentions, as well as the moderating effect of audiences' attitude towards vloggers and brands.

Organization of the Thesis

This thesis consists of five chapters. The first one is an introduction of the research background and methodology. The second chapter is the review of current literature relevant to this study. In Chapter 3, the methodology including research procedure, sample, and measurement are discussed. Chapter 4 presents results of the two studies. Additionally, the final chapter includes the discussion of research findings, the limitation of this thesis, as well as the practical application.

Chapter 2: Literature Review

This chapter detailed the scholarship relevant to this thesis mainly in three aspects. The first section provides a discussion of the research and theories related to brand endorsements. The next section includes a discussion of the context of YouTube influencer marketing and the rise of health-related vlogs. Finally, text and sentiment analysis are also examined in this chapter as one of the main research methods of this thesis.

BRAND ENDORSEMENT AND INFLUENCER AS THE ENDORSER

The marketplace is becoming more user and consumer-led, offering marketers some new challenges (Kapitan & Silvera, 2015). Due to the large amounts of information that bombards consumers online, many prefer to rely on experts or celebrity reviews when researching products, which contributes to the popularity of brand endorsements being utilized as a marketing tactic. With the development of social media, brand endorsements are not only utilized in traditional media but are prominently featured on social media (Munnukka, 2019), Social media influencers have become the main channel for marketers to endorse their products, brands, and services. They are early adopters and opinion leaders of a certain area, who have considerable influence on the target audience demonstrated by a large number of followers. Most of them create their own engaging content, which attracts more loyal audiences than the advertising style made by the brand itself (Newberry, 2019).

For many people, especially among Gen Z, watching a YouTube video or checking the Instagram page of their favorite influencers is a daily routine. Thus, advertising based on word-of-mouth diffusion in social media has become very important in the digital marketing landscape. In that context, the influence and value of social media and social media influencers have become a prominent strategy used in the digital and web marketing industries (Lagrée, 2017). The data shows that in 2019, at least half of US and UK digital marketers spent 10% of their digital marketing budgets on influencer marketing, with the spending showing an increase in 2020 (Newberry, 2019).

Social media marketing and influencer marketing are widely used due to the many advantages they provide to companies. First, influencer marketing brings more loyal customers. Studies about WOM (word-of-mouth) marketing have demonstrated that the repurchase intention and recommendations are much higher in influencer marketing, which proves the loyalty of consumers (Gruen, 2006). Influencers on social media constitute themselves as a brand and followers are influenced by each "promotion" campaign of the influencer. The more they know about the influencer, the more credibility and loyalty the influencer gets.

Besides, the user-generated content created by the influencer appears more trustworthy for consumers. Compared to the advertising made by the brand, influencers always create a scene of their daily routine, which offers a realistic imagination for consumers. The daily scene created by the brand itself is different from the one created by the influencer for the influencer's followers. To some extent, followers believe they are similar to the influencer (Zainab, Zahra & Shilan, 2020), which is why the influencers' promotion content is considered more trustworthy.

EFFECTIVENESS OF BRAND ENDORSEMENT

There are few ways to measure the effectiveness of a video ad. For display ads, a basic and important measure is the ad completion rate. The ad completion rate focuses on the percentage of viewers who watched completely without abandoning in the middle (Krishnan & Sitaraman, 2013). The goal of many ad campaigns is to try to maximize the completion rate.

CTR (click-through-rate) is another metric to measure ad efficiency, which is about the percentage of users who click on a link associated with the ad during or after watching the ad. However, CTRs for video ads are a misleading and sometimes inaccurate measurement that does not account for any emotional engagement with a brand.

Compared to traditional advertising strategies like display ads or offline promotions, the effectiveness of brand endorsements is hard to measure. The previous study shows that audiences' trust in the influencer, the way influencers use and value a brand, and influencers' fitness with the endorsement brand are essential if the effectiveness of brand endorsement is to be achieved (Kapitan & Silvera, 2015).

Influencers' Trustworthiness

Influencers' own connection to the audience is the baseline of the brand endorsement effectiveness. Before the purchase intention is made, consumers' trust in a product's information provider is crucial (Silverman, 2001). For influencers, there are numerous factors impacting trust: credibility, similarity, and familiarity (Zainab, Zahra & Shilan, 2020).

Similarity means more interpersonal attraction and more persuasion power. Audiences are attracted to follow an influencer with similar interests. Compared to top celebrities, social media influencers are more likely to share their daily life and values with the public (Schouten, 2019), making them seem more approachable and trustworthy.

Higher familiarity and credibility make message receivers more comfortable and easier to be persuaded. The interaction between influencers and followers is asymmetrical, while familiarity and credibility could help decrease or even eliminate the untrustworthiness caused by this kind of unequal interaction (Martensen, Brockenhuus-Schack & Zahid, 2018).

How Influencers Show and Value the Endorsement

According to the Federal Trade Commission (FTC) regulations of many social media platforms like YouTube, the product that is sponsored by a brand should be disclosed to the public (*Mediakix*, 2014). This regulation makes the trustworthiness of brand endorsement lower than some native advertisements. However, the content that influencers self-generate can compensate for the deficiencies. Recently, brands have begun partnering more often with micro-influencers, who despite having a lower number of followers, represent niche communities and often have higher engagement on their posts (*Digital Strike*).

With the same budget, micro-influencers could generate a more detailed post or video about the function and experience of products. The professional and analytical expressions could give consumers a better understanding of the product and more likely to make them become loyal consumers.

Influencers' Fitness with the Endorsement Brand

According to Parmentier & Fischer (2012), influencers performing themselves as a personal brand "need to 'fit in' with the expectations of the field before they can 'stand out' from other competitors". Therefore, influencers who are a good match with the endorsed product, service, or brand could foster consumers' positive attitude towards the endorsement brand.

In sum, endorsement effectiveness is measured by the change of the audience's attitude toward the endorsed brand (Munnukka, 2019). Influencer, or influencer marketing, operates as an agency of the endorsed product, and trustworthiness is the key to gaining consumers' positive intentions.

YouTube Influencer Marketing

As the world's largest UGC video website, YouTube has a large influence on various countries, fields, and people. In this regard, it has become an irreplaceable platform for marketers who can use it to address different goals towards different segments (Soukup, 2014). Although the rise of TikTok cannot be ignored in recent years, YouTube remains the primary platform for video influencers, especially for vloggers to post their content.

Parasocial Interaction Between YouTube Influencers and Consumers

The relationship between YouTube influencers and consumers can be explained by the concept of parasocial interaction. Parasocial interaction (PSI) refers to the relationship between the performer in mass media and the audience. In the field of marketing, PSI can be defined as "an illusionary experience, such that consumers interact with personas (i.e., mediated representations of presenters, celebrities, or characters) as if they are present and engaged in a reciprocal relationship" (Horton & Wohl, 1956, p. 135).

Nowadays, PSI is not only strictly used by traditional media channels, but it can also be applied in online environments, especially for social media platforms. For celebrities or brand companies, the PSI on social media could help them develop a strong relationship with consumers that results in loyal purchase intentions (Labrecque, 2014).

Apart from the intelligent machine response system that is widely used by many brands, another common PSI relationship on social media platforms occurs among influencers and their audience. PSI is the key to building the trust of brand endorsement between YouTube influencers and consumers (Chung & Cho, 2017). As mentioned earlier, trustworthiness is essential for the effectiveness of brand endorsement. Compared to traditional mass media, PSI on social media has advantages of interactivity and openness (Labrecque, 2014), especially when influencers are willing to show their personal details (Chung & Cho, 2017). This characteristic is more obvious among vlog influencers since vlogs are about sharing personal details and enhancing open communication (Paula & Antoniya, 2017).

Rise of Vlogs and Health-related Vlogs

Given the topic of this study, it is important to explain why vlogs and health-related vlogs have been chosen as the research objects. Christian defines vlogs as "Video Weblogs, video blogs, or video logs— a specific kind of video on YouTube. A vlog is many things, and different things to different people, but most broadly, it is an expression of a self." (Christian, 2009) The rise of vlogs shows the popularity of real personalities and real-life stories. The interactive functions like comments, likes, and messages on YouTube also help create the community for vlogs' rise (Christian, 2009).

Nowadays, many people are pursuing their personal health goals, including a healthy diet or physical exercise. The current consuming trend of the whole world is the phenomenon of fitness and health (Jesper & Thomas, 2014). The previous study shows that the reason people post their images about healthy activities on social media or follow health-related influencers is to seek social support and provide support for others (Chung, 2017). Among all the various types of vlogs, health-related vlogs have an increasing trend (Paula & Antoniya, 2017). They are becoming a new way to convey healthy lifestyles and provide social support for others.

Audience Participation and Commenting Behavior in YouTube Vlogs

According to the previous study, audience participation refers to how audiences engage in vlogs, including liking, disliking, commenting, sharing, and uploading videos. Besides, reading other comments and viewing the videos are also included in audience participation (Khan, 2017). In another earlier study, the actions that individuals deal with content has been divided into two categories: consumption and participation. Consumption includes watching videos, viewing likes/dislikes and reading comments with no response. Participation entails all user-to-user and user-to-content interaction like commenting, sharing, liking, and disliking (Shao, 2009).

Amongst audience participation actions, the comment is the most complex one. That is because the audience could not only leave comments under vlogs but also read them (Khan, 2017). Therefore, the consequence of reading comments can also be one of the motivations for writing them. The previous study shows that seeking and giving information, as well as social interactions and relaxing entertainment are some motivations of audience participation (Khan, 2017).

The current study focuses on the audience participation mentioned above. With the help of text, sentiment, and network analysis, different aspects of comments will be analyzed and related to vloggers' trustworthiness and product endorsement effectiveness.

TEXT ANALYSIS AND SENTIMENT ANALYSIS OF YOUTUBE COMMENTS

The Current Condition of Text Analysis

Before doing the text analysis of the social media data, the first step is to find a way to collect the necessary information. Felt (2016) mentioned different approaches for data collection, including SVM, some API website tools including Yourtwapperkeeper, NodeXLNetlytic, Storify, and DMI TCAT. Currently, there are problems facing this area: social media platforms increasingly restrict free access to such data, especially through Application Programming Interface (API). Twitter, for example, sells its data to some companies, but these data are only available to limited groups. A more professional way to collect such data from social media platforms is through the use of coding languages like Python. For communication researchers, this is a high-demanding method. Thus, it is now much harder for text analysis researchers to access dataset for their research.

Some researchers use Netlytic to collect content data from social media websites like Twitter, YouTube, and Instagram. Netlytic is used for text and network analysis, including keyword analysis, name network, and chain network analysis. One of its advantages is that it can help with finding online discussions and the corresponding user and community characteristics. In previous studies, researchers used Netlytic to find the author and image characteristics of Instagram. By conducting both text analysis and network analysis, they found the most frequently mentioned keywords inside this hashtag, and the number of nodes and ties inside the community (Santarossa, 2016).

Numerous studies have mentioned the difficulties of social media text analysis (Hussain, 2018; Kabir, 2018; Asghar, 2015). The limit of gathering social media data and the lack of Internet slang dictionaries are the main reasons for the inaccuracy of study results.

LIWC2015 as the Text and Sentiment Analysis Tool

The main text analysis tool that will be used in the current study is LWIC2015. LIWC2015, or Linguistic Inquiry and Word Count 2015, is a text analysis application developed by James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn from The University of Texas at Austin (Pennebaker, 2015). This tool uses the newest dictionary compared to its older versions (LIWC2001 and LIWC2007) and brings a more accurate analysis of cognitive, structural, and emotional variables.

There are few output variables of LWIC2015, and the frequently used variables are the summary language variables, including analytical thinking, clout, authentic, emotional tone, words, and dictionary words. Three of the variables will be used in the current study: analytical thinking, clout, and tone.

The first, analytical thinking, represents the text more in a narrative linguistic style (have more descriptive words) or in an analytical linguistic style (contain more references to objects and concepts) (Aleti, 2019). Previous research (Escalas, 2007) indicates that more narrative expressions could leave more favorable brand impressions.

The second variable is clout, which represents the expressions that are more externally focused (contain more I-words) or internally focused (contain more we-words) (Kacewicz, Pennebaker & Davis, 2013). According to a study about social hierarchies (Kacewicz, Pennebaker & Davis, 2013), an externally focused linguistic style has more power to convince peers. The final variable is tone, which uses sentiment analysis to evaluate the positive, neutral, or negative emotions within context. Here, a study demonstrates that the content with a positive emotion is more likely to be shared (Berger, 2011).

In the current study, the three variables mentioned above will be analyzed and the linguistic style of the health-related vlogs' comments will be one of the research questions.

Text and Sentiment Analysis in YouTube Comments

There are few difficulties in the text analysis of YouTube comments. They are one of the short message formats, like phone texting message, which would lead to some misunderstanding during the text analysis procedure (Walther & Parks, 2002). Recently, some new text analysis applications like SentiWordNet (Esuli, 2014) or LIWC2015 have improved the coverage of text dictionaries and sentiment classification, which increase the accuracy of text and sentiment analysis towards social media platforms.

A previous study found that the sentiment of YouTube comments is related to the number of comments and the view number of the video (Hussain, 2018), and (Kabir, 2018) uses R programming for text analytics and to create a word cloud based on sentiment analysis. Unlike the study about YouTube TED talk comments (Veletsianos, Kimmons, Larsen, Dousay & Lowenthal, 2018), most health-related vlogs feature the same topic, which barely rely upon topical sentiment analysis.

For companies with YouTube accounts or those that do influencer marketing on YouTube, the sentiment of YouTube comments is an easier way to collect consumer opinions and provide insight into future marketing decisions (Younis, 2015).

Chapter 3: Research Questions

Based on the previous theory and study mentioned in the literature review, the following research questions and related variables are conducted.

RESEARCH QUESTIONS

RQ1 What is the linguistic style of health-related YouTube influencers' video comments?

RQ2 What is the main factor that affects the linguistic style of health-related influencer video comments?

RQ3 Is there any relationship between customers' purchase intention about healthy products and their positive commenting intention?

RQ4 Do customers' attitudes towards the healthy product and YouTube influencer moderate the relationship between purchase intention and positive commenting intention?

VARIABLES

Study 1

Dependent variables**:** Tone (negative/positive), Analytical thinking (narrative/analytical), Clout (internally/externally)

Independent variables**:** Number of opinion leaders, Number of comments, Numbers of subscribers, the ratio of likes and dislikes, and whether the video is sponsored or not

Study 2

Dependent variable**:** Purchase intention

Independent variable**:** Positive commenting intention

Moderating variables**:** Product attitude, Influencer attitude

Chapter 4: Methodology

This chapter evaluates how the two studies have been conducted. In Study 1, the research procedure including data collection and text analysis, the measurement of variables, and the statistical strategy are discussed. As for Study 2, it examines the procedure, sample, and measurement of the survey.

STUDY 1

Data Collection

In order to answer RQ1 and RQ2, 60 videos about healthy food, diet, and exercise have been collected from YouTube. According to the Central limit theorem, if the sample size is large (n>=30), the sample mean is normally distributed. Therefore, for the following statistical analysis, 30 sponsored videos and 30 non-sponsored videos have been collected by searching the keywords like "healthy diet", "healthy breakfast", "healthy food" on YouTube.

The following are the criteria for inclusion of the videos for analysis.

First, all videos are published by individual influencers. Since this research is about health-related influencer videos, many related videos published by the official account of some brands and products cannot be collected as objects. Second, to measure one of the independent variables, and whether the video is sponsored or not, both sponsored videos and not-sponsored videos should be taken into account. According to YouTube, every video that is sponsored needs to follow the FTC regulations. YouTubers should disclose the sponsor's messages to audiences at the beginning of the video or include the sentence "This video is sponsored by …" in the "Show More" section under the video. Further, if the video is not sponsored, YouTubers could choose to say nothing or disclose the sentence “This video is NOT sponsored—All opinions are my own.” in the “Show More” section.

Finally, the 60 videos have been collected including two groups, 30 sponsored, and 30 non-sponsored. All the video links are included in Appendix 1. When collecting the videos, the number of influencers’ subscribers and the ratio of likes and dislikes have been recorded. In the next step, network and text analysis tool Netlytic has been used to collect the rest of the independent variables of each video. Netlytic is an API tool that could retrieve data from YouTube, Twitter, Reddit, and other social media platforms, and also do text and network analysis including keyword analysis, name network, and chain network analysis.

After entering the link of a YouTube video, Netlytic collects all comments in text, and do keywords, sentiment, and chain network analysis. The results are provided by CSV files. The files are then downloaded and cleaned, and the last two independent variables, number of opinion leaders and number of comments is then recorded. These will be all the independent variables in the following study.

Text Analysis

Netlytic could yield some basic text analysis results, but this is not shown as numerical numbers, especially for sentiment analysis. Thus, the software called LIWC2015 would be used for text analysis. In this step, the comment text is imported into LIWC2015, and the results have been shown in many categories provided by LIWC2015. To answer the research questions, the results of the following categories have been recorded:

1) tone (negative/positive), 2) analytic (narrative/analytical), 3) clout (internally/externally). These will be the dependent variables in the following study.

Measurement

All the dependent variables have been scaled by 0 points to 100 points according to the output of LIWC2015.

Analytic

The “Analytic” category shows the text is shown in a reasonable, logical, or formal way, contrary to the narrative style, which shows less analytical thinking and is mostly expressed by personal experience. A higher number in this category means there’s more analytical thinking.

Clout

This variable refers to the leadership status among the text. If people have more confident expression, the score of this item will be higher. Since it also shows the interaction with others, it is highly connected with the text community environment, especially when the text data collected from the online social media is analyzed. According to the standard function word dictionaries, a higher number in this category means this comment uses more second-personal singular (you) with the perspective of high expertise. In the field of social media, “Clout” can also be used to identify if the text has an internal or external linguistic style. Basically, a higher number means the statement is more external.

Tone

There are several categories in LIWC2015 to measure the emotional style of the text. First, under the “Summary Language Variables,” the item about emotional tone is abbreviated to “Tone.” Besides, under the “Psychological Processes,” there are also two categories of positive emotion and negative emotion (abbreviated to “posemo” and “negamo”).

Since LIWC2015 mostly focuses on functional words and phrases, the resulting number of “Tone” is always much higher than “posemo” (positive emotion) or “negamo” (negative emotion). Take the output data provided by (Pennebaker, 2015) as an example. The grand mean of "Tone" is 54.22, while the grand means of "posemo" and "negamo" are only 3.67 and 1.84, respectively. According to the scale of LIWC2015 which ranges from 0 points to 100 points, the too low numbers might bring some misunderstanding, especially when the other two dependent variables also have an average number of around 50.

Based on the above situation, "Tone" will be a better choice to measure emotional sentiments compared to "posemo" (positive emotion) or "negamo" (negative emotion). The measurement of "Tone," as shown Pennebaker, 2015, will be as follows: the higher score in "Tone" means there are more positive words in the text. If the number of "Tone" is under 50, there will be more negative emotion than positive emotion.

Analytical Strategy

After getting all the dependent variables and independent variables, the next step is using SPSS to conduct the following tests:

1) Get the descriptive statistic information of all the independent variables,
2) Get the descriptive statistic information of all the dependent variables,
3) Create linear regression models between each independent variable and dependent variables.

The results from the above three steps should give the answer to RQ1 and RQ2.

STUDY 2

Procedure

A survey has been posted on Amazon Mechanical Turk to find a relationship between the customers' purchase intentions towards healthy products and their positive comments' intentions under the health-related influencer videos. The survey consisted of five parts. The first section measures respondents' attitudes toward healthy products. The second section measures respondents' attitudes toward influencers, including similarity, familiarity, and credibility. The third section measures customers' purchase intentions toward healthy products, and the fourth part section measures their commenting behavior under the YouTube influencers' videos. Finally, the respondents' demographic information was recorded, including age, gender, marital status, political affiliation, and race-ethnicity.

The survey consisted of 16 questions, and 197 respondents fully completed the survey, all of which were used in the data analysis.

Sample

There is a total of 197 respondents in the survey, including 126 males and 71 females. Around 88.8% are married and approximately 82.7% live in the U.S. Respondents with ages from 31 to 49 (82,41.6%) and 18 to 30 (74, 37.6%) account for the majority. In terms of education level, the respondents have mainly earned either a bachelor's (132, 67%) or master's degree (54, 27.4%). As for the race of the respondents, the majority were either White (127, 64.5%) or Asian (54, 27.4%).

Measurement

The survey includes items measuring the product attitude, influencer attitude, purchase intentions of healthy products, and the positive commenting intentions under YouTube influencers' videos.

Moderating Variables

There are two moderating variables in this study: respondents' attitude toward healthy products and respondents' attitude toward influencers. The former involves measuring the respondents' attitudes toward healthy products. These questions were measured on a 7-point Likert Scale, from "Strongly Disagree" to "Strongly Agree." There were 5 items on the scale, 2 of which are distracting items. The other 3 items showed a mean of 5.2 and a standard deviation of 1.3.

The second scale measured the respondents' attitudes toward influencers. This scale featured 7 statements to test respondents' influencer attitudes with a 7-point Likert Scale (M = 5.1 and SD = 1.2).

Independent Variable

Respondents' positive commenting intentions were measured in this part. There were 11 statements measuring respondents' commenting behavior, 2 of which are about positive commenting intention. Further, the statements are measured by a 7-point Likert Scale (M = 5.23, SD = 1.4).

Dependent Variable

Purchase intention is the dependent variable in this research and there were 3 items on its scale. These include statements such as "I used to buy the healthy products that the influencer recommends in the video," "I prefer to buy healthy products from an influencer I like," "I prefer to buy healthy products from the influencer whose video has more positive comments." A 7-point Likert Scale has been used (M = 5.4, SD = 1.25).

Extraneous Variables

Demographic questions were created to measure the extraneous variables, including gender, age, marital status, political affiliation, race, place of residence, and education level.

Chapter 5: Results

In this chapter, the findings of the two studies are examined. SPSS is used as the statistic tool to analyze variables by ANOVA and linear regression model.

STUDY 1

The Descriptive Statistics of Independent Variables

SPSS was used to analyze the descriptive statistics of the four independent variables: number of YouTubers' subscribers, the ratio of the videos' likes and dislikes, the number of video comments, and the number of opinion leaders among the comments. The output is illustrated in Table 1.

In total, 50545 comments from 60 YouTube videos have been collected. The range of comment counts in the video was from 34 to 3190, with a mean of 831.93 and a standard deviation of 656.79, showing a positive skew.

When it comes to the number of YouTubers' subscribers, the maximum is 15.5M and the minimum is 1670. This large variance between the number of subscribers gives the research a more representative result, since the sample covered both popular YouTubers and rising YouTubers.

The ratio of likes and dislikes has a range from 16.79 to 263, M = 90.08, SD = 51.94, illustrating a large variation and positive skew.

Statistics

		Ratio	number of subscribers	Comments	Opinion Leaders
N	Valid	60	60	60	60
	Missing	0	0	0	0
Mean		90.0820327	2261730.33	831.93	4.38
Median		79.4571552	998000.000	692.00	4.00
Mode		16.791667[a]	998000.000	34[a]	1
Std. Deviation		51.9395963	3771529.42	656.793	3.484
Variance		2697.722	1.422E+13	431376.470	12.139
Skewness		.970	3.053	1.461	.567
Std. Error of Skewness		.309	.309	.309	.309
Minimum		16.7916667	1670.0000	34	0
Maximum		263.000000	15500000.0	3190	12

Table 1: Descriptive Statistics Analysis of Independent Variables

The Linguistic Style of Health-related Video Comments

To give the solution for RQ1, all 50545 comments were analyzed by LIWC2015 and the output of all three categories, "Analytic," "Clout," and "Tone" have been recorded. SPSS was used to do the descriptive statistics analysis of the three dependent variables (See Table 2).

Statistics

		Analytic	Clout	Tone
N	Valid	50545	50545	50545
	Missing	0	0	0
Mean		45.7156	58.2234	64.8045
Std. Error of Mean		.15944	.15266	.16924
Median		38.6000	50.0000	96.0200
Mode		93.26	50.00	99.00
Std. Deviation		35.84630	34.32192	38.04968
Variance		1284.957	1177.994	1447.778
Range		99.00	99.00	99.00
Minimum		.00	.00	.00
Maximum		99.00	99.00	99.00
Sum		2310693.82	2942899.81	3275543.06

Table 2: Descriptive Statistics Analysis of Dependent Variables

Category	Blogs	Expressive writing	Novels	Natural Speech	NY Times	Twitter	Grand Means	Mean SDs
Linguistic Processes								
Word count (mean)	3206.45	408.94	65716.49	794.17	744.62	660.24	11921.82	10274.32
Analytic	49.89	44.88	70.33	18.43	92.57	61.94	56.34	17.58
Clout	47.87	37.02	75.37	56.27	68.17	63.02	57.95	17.51
Authentic	60.93	76.01	21.56	61.32	24.84	50.39	49.17	20.92
Tone	54.50	38.60	37.06	79.29	43.61	72.24	54.22	23.27

Table 3: LIWC2015 Output Variable Information

The mean "Analytic" of this study is 45.71, showing a significantly lower score than the grand mean (= 56.34) of all kinds of text categories provided by LWIC2015 (See Table 3). The sample taken from the *New York Times* shown in Table 3 has the highest "Analytic" number of 92.57 among all the text categories. This is because of its objective and analytical passages. Since the video samples used in this study are all about healthy products, a healthy diet, or exercise, the analyzed comments tended to compliment the YouTubers, describe viewers' own experiences, and details pertaining to other personal or daily topics. Therefore, the linguistic style of the comments under healthy YouTube influencers' videos is more narrative than analytical.

The average number of "Clout" according to the LWIC2015 handbook is 57.95, while the result of this study shows 58.22. A higher number in this category means the sample comments mostly use second-person singular and first-person plural pronouns and implies a more confident and external expression. YouTubers and the viewers' interaction are unidirectional. Therefore, the viewer is more likely to comment in an external way.

Based on the online community atmosphere of social media, positive and neutral comments account for the largest proportion. The output of text analysis also confirms this result. According to the report from LIWC2015, the average "Tone" of all kinds of text categories is 54.22, while the output from this study shows the average is 64.8. People prefer to leave compliments to the YouTuber, especially under the video that they show a healthy and delicate daily life. Therefore, the comments under health-related YouTube influencers are narrative, external, and positive. RQ1 has been answered.

Linear Regression Models Between Predictors and Dependent Variables

To answer RQ2, linear regression models have been created by SPSS between independent variables and each dependent variable.

Before the linear regression models were created, the variable concerning whether the video is sponsored or not should be recoded because it has multiple categories and is not a numeric variable. The variable "sponsor" was recoded as "sponsored = 0," and "not sponsored = 1."

Analytic

The linear regression analysis (refer to Table 4) showed how each independent variable predicts "Analytic." The results showed that 0.2% of the variance was explained by the five predictors. Whether the video is sponsored or not (β = -.029, p = .000) and number of YouTubers' subscribers (β = -.022, p = .000) have negative impacts on the analytical thinking degree of video comments. This implies that the more subscribers the influencer has, the more narrative linguistic style comments may be found under his/her video. Besides, people are more likely to leave analytical comments under the non-sponsored videos compared to the sponsored videos.

Number of the video comments (β = .029, p = .000) and number of opinion leaders (β = .012, p = .010) have positive impacts on the analytical thinking degree of video comments. This is interpreted to be, the more comments and opinion leaders there are, the more analytical linguistic style the comments are.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.046[a]	.002	.002	35.80948

a. Predictors: (Constant), sponsor, opinion, comments, subscribers, ratio

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45.506	.505		90.078	.000
	subscribers	-1.935E-7	.000	-.022	-4.666	.000
	ratio	-.005	.004	-.006	-1.295	.195
	comments	.001	.000	.029	6.483	.000
	opinion	.118	.046	.012	2.584	.010
	sponsor	-2.093	.335	-.029	-6.251	.000

a. Dependent Variable: Analytic

Table 4: Regression Analysis Between Independent Variables and Analytic

Clout

The linear regression analysis (refer to Table 5) showed how each independent variable predicts "Clout". The results of the regression indicated the five predictors explained 0.1% of the variance. The regression analysis showed that the ratio of likes and dislikes alone marginally predicts the clout of video comments ($\beta = .009$, $p = .058$), which means that the video with a higher ratio of likes and dislikes is more likely to have external comments style.

Meanwhile, the number of subscribers negatively predicts the clout of video comments ($\beta = -.03$, $p = .000$). The more subscribers the influencer have, the more likely the video has internal comments. The variable concerning whether the video is sponsored or not positively predicts the clout of video comments (β = .021, p = .000). Sponsored health-related videos are more likely to receive external comments.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.036[a]	.001	.001	34.30075

a. Predictors: (Constant), sponsor, opinion, comments, subscribers, ratio

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	57.992	.484		119.841	.000
	subscribers	-2.532E-7	.000	-.030	-6.373	.000
	ratio	.007	.003	.009	1.898	.058
	comments	.000	.000	-.006	-1.350	.177
	opinion	-.023	.044	-.002	-.533	.594
	sponsor	1.468	.321	.021	4.579	.000

a. Dependent Variable: Clout

Table 5: Linear Regression Analysis Between Independent Variables and Clout

Tone

According to the result of the regression analysis, all five independent variables significantly predict the emotional style of the comments and explained only 1.1% of the variance (See Table 6).

Number of the video comments (β = -.031, p = .000) and number of YouTubers' subscribers (β = -.058, p = .000) have negative impacts on the sentiment style of video comments. This is interpreted as, the more comments and subscribers there are, the more negative the comments there are. Ratio of likes and dislikes (β = .067, p = .000), number of opinion leaders (β = .033, p = .000) and the whether a video is sponsored or not (β =

.057, p = .000) have positive impacts on the emotional style of video comments. The more ratio and opinion leaders there are, the more positive the comments there are. Besides, sponsored videos are more likely to have positive comments.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.106[a]	.011	.011	37.83804

a. Predictors: (Constant), sponsor, opinion, comments, subscribers, ratio

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	58.896	.534		110.332	.000
	subscribers	-5.487E-7	.000	-.058	-12.519	.000
	ratio	.054	.004	.067	14.014	.000
	comments	-.001	.000	-.031	-6.888	.000
	opinion	.345	.048	.033	7.154	.000
	sponsor	4.389	.354	.057	12.407	.000

a. Dependent Variable: Tone

Table 6: Linear Regression Analysis Between Independent Variables and Tone

RQ2 has been solved.

STUDY 2

Reliability

Before analyzing the data to find the results of the RQ3 and RQ4, the reliability of the items was tested to measure positive commenting intention (IV), purchase intention (DV), product attitude, and influencer attitude (MV). It is important to make sure each variable has a highly reliable scale (Cronbach's Alpha above 0.65). Tables 7- 10 show the reliability of all the four variables separately and the Cronbach's Alphas are all larger than 0.65.

Reliability Statistics

Cronbach's Alpha	N of Items
.894	3

Table 7: Reliability Statistics of Purchase Intention

Reliability Statistics

Cronbach's Alpha	N of Items
.873	3

Table 8: Reliability Statistics of Product Attitude

Reliability Statistics

Cronbach's Alpha	N of Items
.941	7

Table 9: Reliability Statistics of Influencer Attitude

Reliability Statistics

Cronbach's Alpha	N of Items
.800	2

Table 10: Reliability Statistics of Positive Commenting Behavior

Correlation Between Dependent Variable and Independent Variable

The correlation was checked to understand the relationship between positive commenting intentions that respondents were shown and their intentions to purchase the products shown in the influencer's video. After data analysis, a positive, significant correlation was found, r = .88, p = .000 ($p < .05$), which means respondents who have stronger intention to leave positive comments are more likely to purchase the products recommended in the video (Refer to Table 11).

RQ3 has been answered.

Correlations

		PurchaseIntention	Positivecomment
PurchaseIntention	Pearson Correlation	1	.881**
	Sig. (2-tailed)		.000
	N	197	197
Positivecomment	Pearson Correlation	.881**	1
	Sig. (2-tailed)	.000	
	N	197	197

**. Correlation is significant at the 0.01 level (2-tailed).

Table 11: Correlation Between Positive Commenting Intentions and Purchase Intentions

Test Moderating Variables

A test of between-subjects effects was conducted to analyze if respondents' attitudes toward healthy products or attitudes toward the influencer moderates the correlation between positive commenting intentions and purchase intentions.

Table 12 illustrates that respondents' attitudes toward the influencer do not function as a moderating variable (p = .806), while their attitudes toward the product itself have a marginally moderating function (p = .058). Interpreted, this means that the conclusion that respondents' positive commenting intentions increase their purchase intentions does not depend on their attitudes toward the influencer in the video, but marginally depends on their attitudes toward the product itself.

RQ4 has been answered.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.316	.260		8.892	.000
	PK_PC	.098	.052	.921	1.906	.058
	ProductKnowledge	-.458	.283	-.479	-1.620	.107
	IA_PC	-.013	.053	-.117	-.246	.806
	InfluencerAttitude	.571	.291	.563	1.960	.051

a. Dependent Variable: PurchaseIntention

Table 12: Test of Moderating Function of Product Attitude and Influencer Attitude

Chapter 6: Conclusion

DISCUSSION

This study attempts to build a reliable path from the basic audience participation of health-related vlogs to the final purchase intention of the endorsed products. By conducting two studies using of survey and data analysis research methods, the research mainly illustrates two relationships.

The first one as shown in Study 1 demonstrates how audience consumption and participation (liking, disliking, commenting, subscribing) predicts the linguistic style of the video comments. Before the linear regression relationship has been built, the linguistic style of the health-related vlogs' comments is demonstrated as the narratively, externally, and positively focused style. To be specified, this linguistic style combination indicates that the audience of health-related vlogs prefer to mention more personal experiences and feelings using more I-words instead of we-words, and are more likely to give positive feedback to the vloggers.

An asymmetrical relationship is commonly seen in traditional PSI situations (Aleti, 2019), in which the media persona is more powerful and more likely to use externally focused styles. However, the external linguistic style of health-related vlogs' comments implies the weakening of hierarchy and one-sided relationships. As a consequence, this solid PSI fostered between endorsers and the audience could lead to higher brand loyalty and purchase willingness (Chung & Cho, 2017).

Again, audience consumption and participation as predictors have been checked to investigate how they influence the different linguistic styles. The result (Table 13) illustrates that the emotionally focused style is more likely to be influenced. More subscribers and more comments under the vlog indicate there will be less polarized sentiments and more neutral sentiments (Shao, 2009). But the commenting atmosphere under health-related vlogs is mostly positive, which indicates that health-related brands could consider the micro-influencers as their endorser, compared to the top vloggers.

DV/IV	Analytic	Clout	Tone
Subscribers	Negative	N/A	Negative
Endorsement	Non-sponsored	Sponsored	Sponsored
Comments	Positive	N/A	Negative
Opinion Leaders	Positive	N/A	Positive
Ratio of likes and dislikes	N/A	Marginally positive	Positive

Table 13: Summary of Regression Analysis Results in Study

A clear and strong linear regression relationship between positive, emotional comments and the purchase intention has been observed in Study 2. Product review is an important part of word-of-mouth diffusion, and for YouTube vlogs, the messages and emotions conveyed by its comments are quite crucial. In the moderating analysis, the moderating effect of consumers' product attitude is significantly larger than the influencer attitude. This result indicates that for healthy products, consumers care about the product itself more than the endorsers. Brand endorsement through YouTube influencers can be used as a tactic instead of the market share increase strategy.

PRACTICAL IMPLICATIONS

This research can offer some brand endorsement references for health product marketers. With the concept of micro-influencers' rising, many marketers find more options for their endorsers. For the celebrity with mass followers, the brand endorsement may distract the efficiency— the audience is more interested in influencers instead of the brand, and the intrusiveness could also increase.

This thesis could also help YouTube influencers manage their channels and videos. For instance, the study shows that more opinion leaders could lead to more positive comments. Therefore, in order to get more positive feedback, the influencers could try to increase the number of opinion leaders under their videos. As the publisher of the video, the influencers themselves have already been the opinion leaders. The result of network analysis produced by Netlytic shows that if an influencer left comments under his/her own video, the influencer will be the strongest opinion leader. Figure 1 shows the chain network analysis result of one of the sample videos. The account called "pick up limes" is the video publisher and apparently, she received the most replies and became the strongest opinion leader according to the size of the dots.

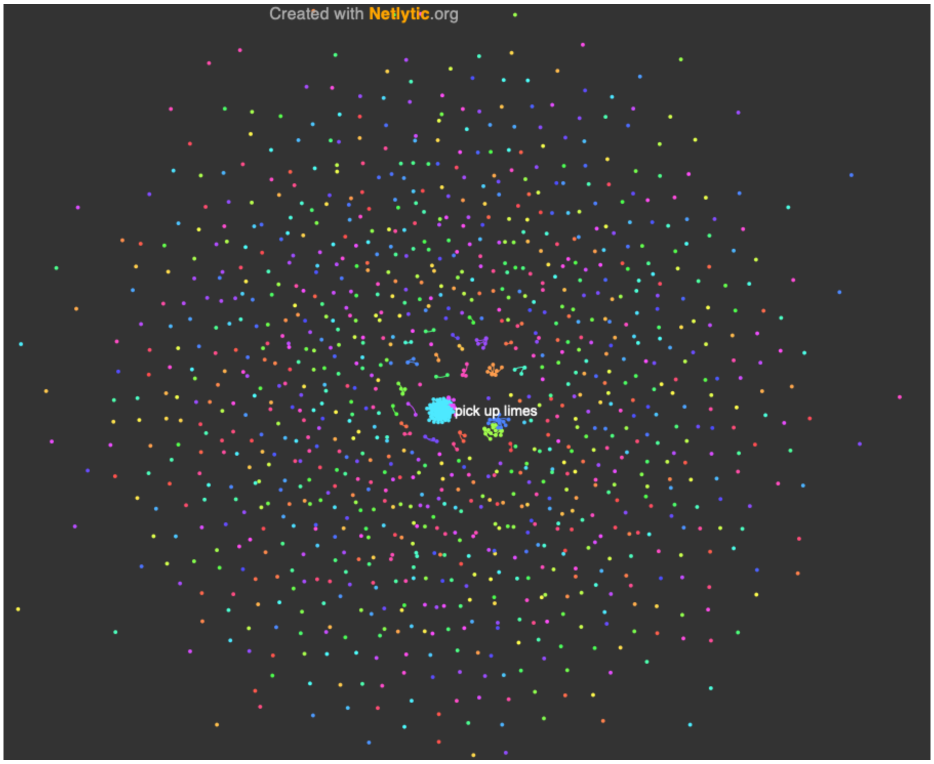

Figure 1: Chain Network Analysis of Video No.6

Therefore, based on the finding above, YouTube influencers could choose to leave their own comment under their videos to get more positive comments. They could also pin their own comments to the top so that the audience could read this at the earliest time. Besides, persuading other popular YouTubers to comment the video is also a great tactic to receive more positive feedback.

LIMITATIONS

There are a few limitations to this thesis. First, it is quite incomplete to only relate the purchase intention with the analysis of comments' sentiments. The sorting and ranking mode of YouTube comments does not very accurately reflect viewers' opinions. Besides, audiences prefer to leave comments under the video they like, compared to the video they are not quite interested in, which also results in the inaccuracy of sentiment analysis.

Second, some variables, including the number of opinion leaders and the ratio of likes and dislikes, are calculated in various ways and could lead to some misunderstanding. Take opinion leaders as an example. Netlytic defines it as the comment that receives more replies. However, according to the default sorting mode of YouTube comments, the comment that receives more replies may not be at the top. For most viewers, the opinion of top comments is more important to them.

Third, it is not quite accurate to use "vloggers" in the thesis since many influencers don't just publish vlogs but make various formats of videos. The concept of vloggers remains fuzzy, and vloggers do not have large differences from normal influencers.

Also in Study 1, the number of R square shows that only a very small part of the dependent variables' variance has been explained by the five predictors (R square = .001, .002, .011). Although some predictors have a significant p-value according to the linear regression analysis, they still have a really small impact on the linguistic style of the comments.

Besides, there are also some flaws in the data analysis process that may result in inaccurate consequences. For example, the main content of collected comments is in English characters. However, there are also many comments that includes emoji or other languages, which brought some difficulties for the text and sentiment analysis. To solve this problem, a more consummate analysis tool is required.

CONCLUSION

Overall, this study finds that the linguistic style of the health-related vlogs' comments is demonstrated as the narratively, externally, and positively focused style. Besides, a positive and significant relationship between positive, emotional comments and the purchase intention has been found. The study also shows that for healthy products, consumers care about the product itself more than the endorsers.

Appendix

Collected YouTube Video Links

Not Sponsored Videos

https://www.youtube.com/watch?v=ylmV10CNJmU

https://www.youtube.com/watch?v=80VyoCETkLQ

https://www.youtube.com/watch?v=FdEQYHGCMeg

https://www.youtube.com/watch?v=HW7iZO0HdLQ

https://www.youtube.com/watch?v=cvnShgiocwQ

https://www.youtube.com/watch?v=mI7WMgsR3WQ

https://www.youtube.com/watch?v=wptpVv2VbYM

https://www.youtube.com/watch?v=xRgWtcFhMeA

https://www.youtube.com/watch?v=DjdbOm6UWgM

https://www.youtube.com/watch?v=bmnh6wTWZXY

https://www.youtube.com/watch?v=4CsntmL50fQ

https://www.youtube.com/watch?v=lCldQpmCVPc

https://www.youtube.com/watch?v=FE2I4igAn34

https://www.youtube.com/watch?v=EVb7LMZap54

https://www.youtube.com/watch?v=LMLSISU0GFc

https://www.youtube.com/watch?v=eQtvmAumVXw

https://www.youtube.com/watch?v=u5px6tWvARA

https://www.youtube.com/watch?v=314fLYhkChk

https://www.youtube.com/watch?v=nvKkQgx0ivc

https://www.youtube.com/watch?v=7eO_t1Y-ofg

https://www.youtube.com/watch?v=4D8KEU3HVyw https://www.youtube.com/watch?v=YZxCc6xMPXI

https://www.youtube.com/watch?v=GohJ60SVP7M

https://www.youtube.com/watch?v=LD3UozF8Hn0

https://www.youtube.com/watch?v=rmwsR8aRXgw

https://www.youtube.com/watch?v=IL5sV3va1yg

https://www.youtube.com/watch?v=ASqMlF7K8_w

https://www.youtube.com/watch?v=2kCOXLkbsKw

https://www.youtube.com/watch?v=qi-TUigDdsI

https://www.youtube.com/watch?v=fmBMXwApMqg

https://www.youtube.com/watch?v=w733i9Umcn0

https://www.youtube.com/watch?v=H-6Qm9KPcmA

https://www.youtube.com/watch?v=agUEe-VTDbs

https://www.youtube.com/watch?v=eu-HNcRyMf4

https://www.youtube.com/watch?v=uxVlinnmaB8

https://www.youtube.com/watch?v=aIlj2O0eGQk

https://www.youtube.com/watch?v=KmCNwbEZm7g

https://www.youtube.com/watch?v=ZV0vafQ_l7Q

Sponsored Videos

https://www.youtube.com/watch?v=FTMCpop3ORA

https://www.youtube.com/watch?v=djImT8VGQxE

https://www.youtube.com/watch?v=ip9tOazjcEU

https://www.youtube.com/watch?v=L9VBoIK3Chw

https://www.youtube.com/watch?v=oNt8bo_PsB4

https://www.youtube.com/watch?v=1_LUmwXwzFY https://www.youtube.com/watch?v=NCmIIx9dUNw

https://www.youtube.com/watch?v=TxGnawb13_0

https://www.youtube.com/watch?v=3KJI9Pagos4

https://www.youtube.com/watch?v=k6Z1BKqRqFs

https://www.youtube.com/watch?v=WMQkE_FlNag

https://www.youtube.com/watch?v=ii2yK2m1JyA

https://www.youtube.com/watch?v=44zZLf4oPh4

https://www.youtube.com/watch?v=lOFHyiSkFkU

https://www.youtube.com/watch?v=MBGH_2pzhig

https://www.youtube.com/watch?v=0zdE5rJ9k4E

https://www.youtube.com/watch?v=Jr0uY2QPkUQ

https://www.youtube.com/watch?v=1PQy_cJV0qU

https://www.youtube.com/watch?v=oI3WZhJ4F3c

https://www.youtube.com/watch?v=TqfwLvOOM7s

https://www.youtube.com/watch?v=CB0bqOCb1RE

https://www.youtube.com/watch?v=s3b0Q8kXI7U

https://www.youtube.com/watch?v=K26QQopxvW0

https://www.youtube.com/watch?v=1VOMWyvGM3k

https://www.youtube.com/watch?v=bj2PXI2hwZs

https://www.youtube.com/watch?v=e0FydG9Aw-A

https://www.youtube.com/watch?v=sLn-t0uhwuo

https://www.youtube.com/watch?v=fYH5qM-bdFc

https://www.youtube.com/watch?v=XcZG_7O-fQs

https://www.youtube.com/watch?v=q65enzklzqo

SURVEY

Section 1 Introduction

Welcome to my survey. This survey is an integral part of my Master thesis at the University of Texas. This survey is completely confidential, none of your information will be linked to your responses. The survey should take about 5 minutes to complete. If at any time you wish to stop participating, simply close the browser. I'm grateful for anyone who decides to participate. Thank you for your support.

Section 2 Product Attitude

Q1 Have you tried any healthy products? (like nutrition powder, drinks, food, etc)

( ) Yes

( ) No

Q2 Where did you get the acknowledgment of these healthy products? Please select all that apply.

[ ] Professional articles or classes

[ ] Gym instructors or friends

[ ] Internet News

[ ] Related influencers/celebrities

[ ] Other: ____

Q3 What's the most decisive factor that makes you buy healthy products?

( ) Online product reviews

( ) Recommendation from friends or the fitness trainer

( ) Recommendation from influencers/celebrities

( ) Product ads

( ) Others: ____

Q4 Choose your level of agreement with each of the following items.

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
I always visit the healthy product website.	○	○	○	○	○	○	○
I have a full acknowledgment of healthy products.	○	○	○	○	○	○	○
The healthy product and website truly helped me improve my physical and motion conditions.	○	○	○	○	○	○	○
I choose to repurchase the product if it helped me improve my physical and motion conditions.	○	○	○	○	○	○	○
I purchase the healthy product regularly.	○	○	○	○	○	○	○
I always try new healthy products.	○	○	○	○	○	○	○

Section 3 Influencer Attitude

Q5 Choose your level of agreement with each of the following items.

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
I'm following influencers on YouTube.	○	○	○	○	○	○	○
I have a lot in common with the influencer I follow.	○	○	○	○	○	○	○
The influencer and I use the same product (we have the same taste in products).	○	○	○	○	○	○	○
I have followed some influencers for a long time. (more than a year)	○	○	○	○	○	○	○
I relate to the influencer on a personal level.	○	○	○	○	○	○	○
I perceive the influencer credible when their message in their post (on social media) is clear.	○	○	○	○	○	○	○
I perceive the influencer as credible when they work with only one brand.	○	○	○	○	○	○	○
I used to watch the influencer videos about healthy food/diet on YouTube.	○	○	○	○	○	○	○

I won't decrease my evaluation of the influencer if the video is sponsored by some healthy products.	○	○	○	○	○	○	○

Section 4 Purchase Intention

Q6 Choose your level of agreement with each of the following items.

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
I would like to buy the healthy products that the influencer recommends in the video.	○	○	○	○	○	○	○
I prefer to buy healthy products from an influencer I like.	○	○	○	○	○	○	○
I prefer to buy healthy products from the influencer whose video has more positive comments.	○	○	○	○	○	○	○

Section 5 Commenting Behavior

Q7 Choose your level of agreement with each of the following items.

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
I would like to leave comments under the influencer videos.	○	○	○	○	○	○	○
I prefer to leave positive comments under the influencer videos.	○	○	○	○	○	○	○
I prefer to leave comments under the video that is not be sponsored by healthy products.	○	○	○	○	○	○	○
I prefer to leave comments under the video that is sponsored by healthy products.	○	○	○	○	○	○	○
I prefer to leave comments under the video sponsored by the products or brands I like.	○	○	○	○	○	○	○
I'll leave comments if I tried the products mentioned in the video.	○	○	○	○	○	○	○
I'll leave comments if I want to try the products mentioned in the video.	○	○	○	○	○	○	○

I prefer to comment under the influencer's video who has more subscribers.	○	○	○	○	○	○	○
I prefer to comment under the influencer's video whose video has more comments.	○	○	○	○	○	○	○
I prefer to comment under the influencer's video whose video has more likes.	○	○	○	○	○	○	○
I prefer to comment under the influencer's video whose video has more positive comments.	○	○	○	○	○	○	○

Q8 What kinds of comments you will leave if you buy the product that is recommended by the video?

[ ] Feedback of the product

[ ] Compliment to the YouTuber

[ ] Reply to other comments that you agree with

[ ] Advice or suggestions for the YouTuber

[ ] Others: ____

Section 6 Demographics

Q9 Gender

( ) Male

( ) Female

( ) Prefer to not answer

Q10 Age

( ) Under 18

( ) 18-30

( ) 31-49

( ) 50-69

( ) Over 70

Q11 What is your marital status?

( ) Single

( ) Married

( ) Civil Union

( ) Divorced

( ) Widowed

Q12 Political Affiliation

( ) Republican

( ) Democrat

( ) Independent

( ) Prefer to not answer

Q13 Race/Ethnicity

( ) American Indian or Alaska Native

( ) Asian

( ) Hispanic or Latinx

( ) Black or African American

( ) White

( ) Native Hawaiian or Pacific Islander

( ) Other

( ) Prefer to not answer

Q14 Do you live in the United States?

( ) Yes

( ) No

Display Question 15:

If Do you live in the United States? = Yes

Q15 Select the state you are from.

Q16 What country are you from?

Q17 Highest level of education completed:

( ) Some High School

( ) High School

( ) Associate's

( ) Bachelor's

( ) Master's

( ) Doctorate

( ) Prefer not to answer

References

Aleti, T. (2019, November). Tweeting with the Stars- Automated Text Analysis of the Effect of Celebrity Social Media Communications on Consumer Word of Mouth. *Journal of Interactive Marketing*, 48, 17-32. Retrieved from https://www.sciencedirect.com/science/article/pii/S1094996819300556

Asghar, MZ. (2015, November). Sentiment Analysis on YouTube: A Brief Survey. *MAGNT Research Report*, 3(1), 1250-1257. Retrieved from https://arxiv.org/abs/1511.09142

Berger, J. (2011, June). Arousal Increases Social Transmission of Information. *Psychological Science*, 22(7), 891-893. Retrieved from https://journals.sagepub.com/doi/10.1177/0956797611413294

Callaghan, D. (2012, January). Network Analysis of Recurring YouTube Spam Campaigns. Retrieved from https://arxiv.org/abs/1201.3783

Christian, A. (2009, October). Real vlogs: The rules and meanings of online personal videos. Retrieved from https://firstmonday.org/ojs/index.php/fm/article/view/2699

Chung, CF. (2017). When Personal Tracking Becomes Social: Examining the Use of Instagram for Healthy Eating. Retrieved from https://dl.acm.org/doi/abs/10.1145/3025453.3025747

Chung, S., Cho H. (2017, March). Fostering Parasocial Relationships with Celebrities on Social Media: Implications for Celebrity Endorsement. *Psychology Marketing*, 34(4), 481-495. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/mar.21001

Does YouTubers "Includes Paid Promotion" tag satisfy FTC guidelines? (2014, January). *Mediakix*. Retrieved from https://mediakix.com/blog/youtube-includes-paid-promotion-tag-ftc/

Escalas, JE. (2007, March). Self-Referencing and Persuasion: Narrative Transportation versus Analytical Elaboration. *Journal of Consumer Research*, 33(4), 421–429. Retrieved from https://academic.oup.com/jcr/article-abstract/33/4/421/1790292

Esuli, A. (2013, June). The User Feedback on SentiWordNet. Retrieved from https://arxiv.org/abs/1306.1343

Felt, M. (2016, April). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society*, 3(1). Retrieved from https://journals.sagepub.com/doi/full/10.1177/2053951716645828

Gruen, T. (2006, April). eWOM: The impact of customer-to-customer online know-how exchange on customer value and loyalty. *Journal of Business Research,* 59(4), 449-456. Retrieved from https://www.sciencedirect.com/science/article/pii/S0148296305001517

Horton, D., Wohl, R. (1956). Mass communication and para-social interaction: Observation on intimacy at a distance. *Psychiatry*. 19 (3), 215–229.

Hussain, MN. (2018). Understanding Digital Ethnography: Socio-computational Analysis of Trending YouTube Videos. Retrieved from http://sbp-brims.org/2018/proceedings/papers/latebreaking_papers/LB_14.pdf

Interpreting LIWC Output. Retrieved from https://liwc.wpengine.com/interpreting-liwc-output/

Jesper A., Thomas, J. (2014). The Fitness Revolution: Historical Transformations in the Global Gym and Fitness Culture. *Sport Science Review*, XXIII(3-4), 91-112. Retrieved from https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A742709&dswid=311

Kabir, AI. (2018). The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R. *Informatica Economică*, 22(1), 25-38. Retrieved from https://www.researchgate.net/profile/Ahmed_Imran_Kabir2/publication/324201476

Kacewicz, E., Pennebaker, JW., Davis, M. (2013, September). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology*, 33(2), 125-143. Retrieved from https://journals.sagepub.com/doi/10.1177/0261927X13502654

Kapitan, S., Silvera, D. (2015, March). From digital media influencers to celebrity endorsers: attributions drive endorser effectiveness. *Mark Letters*, 27, 553–567. Retrieved from https://link.springer.com/article/10.1007/s11002-015-9363-0

Khan, ML. (2017, January). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66, 236-247. Retrieved from https://www.sciencedirect.com/science/article/pii/S0747563216306513

Krishnan, SS., Sitaraman, RK. (2013, October). Understanding the effectiveness of video ads: a measurement study. Retrieved from https://dl.acm.org/doi/10.1145/2504730.2504748

Labrecque, L. (2014, May). Fostering Consumer–Brand Relationships in Social Media Environments: The Role of Parasocial Interaction. *Journal of Interactive Marketing*, 28(2), 134-148. Retrieved from https://www.sciencedirect.com/science/article/pii/S1094996813000650

Lagrée, P. (2017, November). Effective Large-Scale Online Influence Maximization. Retrieved from https://ieeexplore.ieee.org/document/8215581/similar#similar

Lim, Y., Chung, Y., Weaver, P. (2012, July). The impact of social media on destination branding: Consumer-generated videos versus destination marketer-generated videos. *Journal of Vacation Marketing*, 18(3), 197-206. Retrieved from https://journals.sagepub.com/doi/10.1177/1356766712449366

Loren, T. (2016, September). How to Grow Your Business with Instagram Influencer Marketing. Retrieved from https://later.com/blog/instagram-influencer-marketing-for-business/

Martensen, A., Brockenhuus-Schack, S. and Zahid, A.L. (2018, July). How citizen influencers persuade their followers. *Journal of Fashion Marketing and Management*, 22(3), 335-353. Retrieved from https://www.emerald.com/insight/content/doi/10.1108/JFMM-09-2017-0095/full/html

Munnukka, J. (2019, April). “Thanks for watching”. The effectiveness of YouTube vlog endorsements. *Computers in Human Behavior*, 93, 226-234. Retrieved from https://www.sciencedirect.com/science/article/pii/S0747563218306009

Newberry, C. (2019, May). Influencer Marketing Guide: How to Work With Social Media Influencers. *Hootsuite*. Retrieved from https://blog.hootsuite.com/influencer-marketing/

Parmentier, MA., Fischer, E. (2012, July). Positioning person brands in established organizational fields. *Journal of the Academy of Marketing Science*, 41, 373–387. Retrieved from https://link.springer.com/article/10.1007%2Fs11747-012-0309-2

Paula, K., Antoniya, P. (2017). YouTube influence on Well-being brands. Retrieved from https://lup.lub.lu.se/student-papers/search/publication/8914731

Pennebaker, JW. (2015, September). The Development and Psychometric Properties of LIWC2015. Retrieved from https://repositories.lib.utexas.edu/handle/2152/31333

Rogers, EM. (2003, August). Diffusion of Innovations, 5th Edition.

Santarossa, S. (2016, November). #fitspo on Instagram: A mixed-methods approach using Netlytic and photo analysis, uncovering the online discussion and author/image characteristics. *Journal of Health Psychology*, 24(3), 376-385. Retrieved from https://journals.sagepub.com/doi/full/10.1177/1359105316676334

Schouten, A. (2019, July). Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit. Retrieved from https://www.tandfonline.com/doi/full/10.1080/02650487.2019.1634898

Shao, G. (2009, January). Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet Research*, 19(1). Retrieved from https://www.emerald.com/insight/content/doi/10.1108/10662240910927795/full/html

Silverman, B. (2001, November). Implications of buyer decision theory for design of e-commerce websites. *International Journal of Advertising*, 258-281. Retrieved from https://www.sciencedirect.com/science/article/pii/S1071581901905002

Soukup, P. (2014). Looking at, with, and through YouTube. *Communication Research Trends*, 33(3), 3-34. Retrieved from https://search.proquest.com/openview/7d16d19796a46e2fa5794221c9925996/1?pq-origsite=gscholar&cbl=1576344

The New Celebrity Endorsements: How social media influencers change the game. *Digital Strike*. Retrieved from https://www.digitalstrike.com/blog/social-media-influencers/

Veletsianos G., Kimmons R., Larsen R., Dousay TA., Lowenthal PR. (2018). Public comment sentiment on educational videos- Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. *PLoS ONE*, 13(6), e0197331. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197331

Walther, JB., Parks, MR. (2002). Cues filtered out, cues filtered in. Retrieved from https://www.researchgate.net/publication/239489124

Younis, EMG. (2015, February). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. *International Journal of Computer Applications*, 112(5), 0975 – 8887. Retrieved from https://www.researchgate.net/profile/Eman_Younis/publication/272463313

Zainab, A., Zahra, A., Shilan, R. (2020, September). Similarity, Familiarity, and Credibility in influencers and their impact on purchasing intention. Retrieved from https://www.diva-portal.org/smash/get/diva2:1437746/FULLTEXT01.pdf

Zote, J. (2020, January). 40 YouTube stats and facts to power your 2020 marketing strategy. Retrieved from https://sproutsocial.com/insights/youtube-stats/

Vita

Zihan Yang is born and grew up in Henan, China. She completed her high school study at Zhengzhou Foreign Language School in 2014 and entered the Communication University of China, Beijing. She obtained the Bachelor of Arts in Advertising in June 2018. After graduation, she entered the Graduate School at the University of Texas at Austin in August 2018.

Email: zhyang@utexas.edu

This Thesis was typed by Zihan Yang.